

DECLARATION OF NATHANIEL E FRANK-WHITE

1. I am a Records Request Processor at the Internet Archive. I make this declaration of my own personal knowledge.
2. The Internet Archive is a website that provides access to a digital library of Internet sites and other cultural artifacts in digital form. Like a paper library, we provide free access to researchers, historians, scholars, and the general public. The Internet Archive has partnered with and receives support from various institutions, including the Library of Congress.
3. The Internet Archive has created a service known as the Wayback Machine. The Wayback Machine makes it possible to browse more than 450 billion pages stored in the Internet Archive's web archive. Visitors to the Wayback Machine can search archives by URL (i.e., a website address). If archived records for a URL are available, the visitor will be presented with a display of available dates. The visitor may select one of those dates, and begin browsing an archived version of the Web. Links on archived files in the Wayback Machine point to other archived files (whether HTML pages or other file types), if any are found for the URL indicated by a given link. For instance, the Wayback Machine is designed such that when a visitor clicks on a hyperlink on an archived page that points to another URL, the visitor will be served the archived file found for the hyperlink's URL with the closest available date to the initial file containing the hyperlink.
4. The archived data made viewable and browseable by the Wayback Machine is obtained by use of web archiving software that automatically stores copies of files available via the Internet, each file preserved as it existed at a particular point in time.
5. The Internet Archive assigns a URL on its site to the archived files in the format `http://web.archive.org/web/[Year in yyyy][Month in mm][Day in dd][Time code in hh:mm:ss]/[Archived URL]` aka an "extended URL". Thus, the extended URL `http://web.archive.org/web/19970126045828/http://www.archive.org/` would be the URL for the record of the Internet Archive home page HTML file (`http://www.archive.org/`) archived on January 26, 1997 at 4:58 a.m. and 28 seconds (1997/01/26 at 04:58:28). The date indicated by an extended URL applies to a preserved instance of a file for a given URL, but not necessarily to any other files linked therein. Thus, in the case of a page constituted by a primary HTML file and other separate files (e.g., files with images, audio, multimedia, design elements, or other embedded content) linked within that primary HTML file, the primary HTML file and the other files will each have their own respective extended URLs and may not have been archived on the same dates.
6. Attached hereto as Exhibit A are true and accurate copies of screenshots of the Internet Archive's records of the archived files for the URLs and the dates specified in the attached coversheet of each printout.



7. I declare under penalty of perjury that the foregoing is true and correct.

DATE: July 19, 2022

Nathaniel E Frank-White
Nathaniel E Frank-White

EXHIBIT A

<https://web.archive.org/web/20020605231041/http://www.xav.com/scripts/search/help/1048.html>

[[Help: Table of Contents](#)]

How to prevent sections of your pages from being indexed

This help topic describes how to prevent *sections* of a document from being indexed. To prevent an *entire* document from being indexed, see [How to prevent your pages from being indexed](#).

FDSE supports the proprietary "robots" comment tag. This tag allows a web author to apply robots exclusion rules to arbitrary sections of a document. The tag has one attribute, content, with the following possible values:

- noindex - the text enclosed in the tag is not saved in the index
- nofollow - links are not extracted from the text enclosed
- none - enclosed text is not indexed nor searched for links

Values "index", "follow", and "all" are also valid. In practice they are ignored since they are the unspoken defaults.

This feature is expected to fit the customer need of preventing certain parts of a document - such as a navigational sidebar - from being included in the search.

Example:

```
<HTML>
<BODY>

  This text will be indexed.
  <A HREF="foo.html"> this link will be followed </A>

  <!-- robots content="none" -->

    This text will NOT be indexed.
    <A HREF="bar.html"> this link will NOT be followed </A>

  <!-- /robots -->

  <!-- robots content="noindex" -->

    This text will NOT be indexed.
    <A HREF="bar1.html"> this link WILL be followed </A>

  <!-- /robots -->

  <!-- robots content="nofollow" -->

    This text WILL be indexed.
    <A HREF="bar1.html"> this link will NOT be followed </A>

  <!-- /robots -->

  la la la

</BODY>
</HTML>
```

For the example of a navigational sidebar, the "noindex" value would be the best choice.

This syntax was designed to match the robots META tag.

For documents which have both the "robots" META tag and the "robots" comment tag, the most restrictive interpretation will be made, always erring on the side of not indexing or not following.

This tag is *always* respected by FDSE, even when "Crawler: Rogue" is set to 1 and all other robots exclusion rules are ignored. This is because the robots comment tag is proprietary and thus is expected to only be used by the FDSE administrator.

Coding:

All search engines are encouraged to support this syntax, since it results in small index files, faster searches, and more accurate results, and a standard syntax for authors to use. Currently Darryl Burgdorf's [WebSearch](#) supports this syntax too. I've asked Atomz, Google, and Altavista to support it and have received some favorable feedback but I'm not aware of any code changes on their part.

In Perl, the following regular expression will extract the robots sections labeled "none". This regex is designed to be insensitive to whitespace, line breaks, case, and the quoted-vs-bare nature of the attribute value. This regex should be executed as soon as the file is read:

```
$_ = $text;
s'<!--\s*robots\s+content\s*=\s*"none"?|'\s*-->.*?<!--\s*/robots\s*-->' 'isg;
```

For subroutines which extract links, this code will extract the "nofollow" sections:

```
s'<!--\s*robots\s+content\s*=\s*"nofollow"?|'\s*-->.*?<!--\s*/robots\s*-->' 'isg;
```

A final substitution against the "noindex" portion of the file is required before indexing or searching for keywords. For those search engines which just read in text and index it, without extracting links, the more efficient single regex will work:

```
s'<!--\s*robots\s+content\s*=\s*"?(none|noindex)"?\s*-->.*?<!--\s*/robots\s*-->' 'isg;
```

History: originally introduced as the FDSE:ROBOTS HTML tag in version 2.0.0.0031.

As of version 2.0.0.0044, the HTML comment equivalent was introduced. This equivalent format is preferred because it will not cause problems with editors that only support strictly valid HTML, and it will not cause problems with nesting, and it can be adopted as a vendor-independent standard. The equivalent forms are:

```
<FDSE:ROBOTS VALUE="none"> xxx </FDSE:ROBOTS>
<!-- robots content="none" --> xxx <!-- /robots -->

<FDSE:ROBOTS VALUE="nofollow"> xxx </FDSE:ROBOTS>
<!-- robots content="nofollow" --> xxx <!-- /robots -->

<FDSE:ROBOTS VALUE="noindex"> xxx </FDSE:ROBOTS>
<!-- robots content="noindex" --> xxx <!-- /robots -->
```

Handling of these tags is case insensitive and is insensitive to linear and vertical whitespace.

Bugs / reverse compatibility: In early versions, FDSE used the singular format "FDSE:ROBOT" and "!-- robot content". It was later realized that for compatibility with the "robots" META tag and the "robots.txt" file, the plural format would have to be used. Thus, as of version 2.0.0.0048, the FDSE software supports both singular and plural forms, and the documentation promotes the plural form. If you have an earlier version of FDSE software, then you should use the singular form "robot" rather than "robots" in creating your tags. Sorry.

[[Help: Table of Contents](#)]

<https://web.archive.org/web/20060105100607/http://www.xav.com:80/scripts/search/help/1048.html>

[Home](#)

Main Products:

- [FDSE](#)
- [AXS](#)
- [Guardian](#)

[Forum](#)

[Contact Info](#)

[About](#)

[Site Map](#)

©2006 Zoltan Milosevic.
 All Rights Reserved.

[Home](#) > [Fluid Dynamics Search Engine](#) > [Help](#) > **1048**

How to prevent sections of your pages from being indexed

This help topic describes how to prevent *sections* of a document from being indexed. To prevent an *entire* document from being indexed, see [How to prevent your pages from being indexed](#).

FDSE supports the proprietary "robots" comment tag. This tag allows a web author to apply robots exclusion rules to arbitrary sections of a document. The tag has one attribute, content, with the following possible values:

- noindex - the text enclosed in the tag is not saved in the index
- nofollow - links are not extracted from the text enclosed
- none - enclosed text is not indexed nor searched for links

Values "index", "follow", and "all" are also valid. In practice they are ignored since they are the unspoken defaults.

This feature is expected to fit the customer need of preventing certain parts of a document - such as a navigational sidebar - from being included in the search.

Example:

```
<HTML>
<BODY>

  This text will be indexed.
  <a href="foo.html"> this link will be followed </A>

  <!-- robots content="none" -->

    This text will NOT be indexed.
    <a href="bar.html"> this link will NOT be followed </A>

  <!-- /robots -->

  <!-- robots content="noindex" -->

    This text will NOT be indexed.
    <a href="bar1.html"> this link WILL be followed </A>

  <!-- /robots -->

  <!-- robots content="nofollow" -->

    This text WILL be indexed.
    <a href="bar1.html"> this link will NOT be followed </A>

  <!-- /robots -->

  la la la

</BODY>
</HTML>
```

For the example of a navigational sidebar, the "noindex" value would be the best choice.

This syntax was designed to match the robots META tag.

For documents which have both the "robots" META tag and the "robots" comment tag, the most restrictive interpretation will be made, always erring on the side of not indexing or not following.

This tag is *always* respected by FDSE, even when "Crawler: Rogue" is set to 1 and all other robots exclusion rules are ignored. This is because the robots comment tag is proprietary and thus is expected to only be used by the FDSE administrator.

Coding:

All search engines are encouraged to support this syntax, since it results in small index files, faster searches, and more accurate results, and a standard syntax for authors to use. Currently Darryl Burgdorf's [WebSearch](#) supports this syntax too. I've asked Atomz, Google, and Altavista to support it and have received some favorable feedback but I'm not aware of any code changes on their part.

In Perl, the following regular expression will extract the robots sections labeled "none". This regex is designed to be insensitive to whitespace, line breaks, case, and the quoted-vs-bare nature of the attribute value. This regex should be executed as soon as the file is read:

```
$_ = $text;
s'<!--\s*robots\s+content\s*=\s*"?none"? \s*-->.*?<!--\s*/robots\s*-->' 'isg;
```

For subroutines which extract links, this code will extract the "nofollow" sections:

```
s'<!--\s*robots\s+content\s*=\s*"?nofollow"? \s*-->.*?<!--\s*/robots\s*-->' 'isg;
```

A final substitution against the "noindex" portion of the file is required before indexing or searching for keywords. For those search engines which just read in text and index it, without extracting links, the more efficient single regex will work:

```
s'<!--\s*robots\s+content\s*=\s*"?(none|noindex)"? \s*-->.*?<!--\s*/robots\s*-->' 'isg;
```

See also [Compatibility with Atomz.com noindex tag](#).

History: originally introduced as the FDSE:ROBOTS HTML tag in version 2.0.0.0031.

As of version 2.0.0.0044, the HTML comment equivalent was introduced. This equivalent format is preferred because it will not cause problems with editors that only support strictly valid HTML, and it will not cause problems with nesting, and it can be adopted as a vendor-independent standard. The equivalent forms are:

```
<FDSE:ROBOTS value="none"> xxx </FDSE:ROBOTS>
<!-- robots content="none" --> xxx <!-- /robots -->
```

```
<FDSE:ROBOTS value="nofollow"> xxx </FDSE:ROBOTS>
<!-- robots content="nofollow" --> xxx <!-- /robots -->

<FDSE:ROBOTS value="noindex"> xxx </FDSE:ROBOTS>
<!-- robots content="noindex" --> xxx <!-- /robots -->
```

Handling of these tags is case insensitive and is insensitive to linear and vertical whitespace.

Bugs / reverse compatibility: In early versions, FDSE used the singular format "FDSE:ROBOT" and "!-- robot content". It was later realized that for compatibility with the "robots" META tag and the "robots.txt" file, the plural format would have to be used. Thus, as of version 2.0.0.0048, the FDSE software supports both singular and plural forms, and the documentation promotes the plural form. If you have an earlier version of FDSE software, then you should use the singular form "robot" rather than "robots" in creating your tags. Sorry.

"How to prevent sections of your pages from being indexed"
<http://www.xav.com/scripts/search/help/1048.html>

<https://web.archive.org/web/20021125020514/http://www.fusionbot.com/faqs/faq.asp?faq=26>

Can I control the elements that are displayed as part of my search results context?

Title The clickable link, extracted from your page's <TITLE> tag, for each listing returned on your search results.

Description	Displayed below the Title on your search results. Provides a more detailed description of the content on the matching page. This is the text extracted from the <META Description> tag, when present, on your indexed page.
--------------------	--

Body Displayed below the Description on your search results. Provides an excerpt from within the matching document where the actual keyword "match" was found.

Rank The Rank value (percent likelihood of a good match) for each result returned.

URL The URL / link of the matching page.

Size The size of the indexed document, in Bytes.

Date The date the document was last modified and / or created.

The above elements make up the context of your search results.

With the exception of the **Title**, which must be displayed, FusionBot allows you turn on / off any of the additional elements described above, preventing them from being returned within the context of your search results.

To choose which elements you would like to include / exclude from your search results, [login](#) to your account, and click on the **Customization** tab. From there, click the **'results context'** option, and choose which elements you would like displayed. By default, all elements are checked expect the **Rank Value** option, and are therefore included as part of your output.

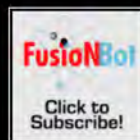
If you did not find an answer to your question, please return to the [Support Page](#) and fill out the form provided.

[Back to FAO List](#)

[Next FAO >>](#)



Featured Customers:



<https://web.archive.org/web/20030905065727/http://www.fusionbot.com/faqs/faq26.asp?pp=&pos=>

INTERNET ARCHIVE

WayBackMachine

1 capture

05 Sep 2003

http://www.fusionbot.com/faqs/faq26.asp?pp=&pos=

Go

AUG

SEP

OCT

05

2002

2003

2004

About this capture


Harness the power of seamless site search solutions

Search our Site: All of FusionBot

LOG-IN | SUBSCRIBE | PACKAGES | FEATURES | CUSTOMERS | ABOUT US | SUPPORT

Frequently Asked Question #26

Can I control the elements that are displayed as part of my search results context?

Your default search results return seven basic elements within the text content, in the following order:

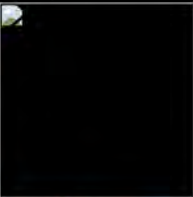
- Title** The clickable link, extracted from your page's **<TITLE>** tag, for each listing returned on your search results.
- Description** Displayed below the Title on your search results. Provides a more detailed description of the content on the matching page. This is the text extracted from the **<META Description>** tag, when present, on your indexed page.
- Body** Displayed below the Description on your search results. Provides an excerpt from within the matching document where the actual keyword "match" was found.
- Rank** The Rank value (percent likelihood of a good match) for each result returned.
- URL** The URL / link of the matching page.
- Size** The size of the indexed document, in Bytes.
- Date** The date the document was last modified and / or created.

The above elements make up the context of your search results.

With the exception of the **Title**, which must be displayed, FusionBot allows you turn on / off any of the additional elements described above, preventing them from being returned within the context of your search results.

To choose which elements you would like to include / exclude from your search results, [login](#) to your account, and click on the **Customization** tab. From there, click the **'results context'** option, and choose which elements you would like displayed. By default, all elements are checked expect the **Rank Value** option, and are therefore included as part of your output.

If you did not find an answer to your question, please return to the [Support Page](#) and fill out the form provided.




Click to Subscribe!

Copyright 1998 - 2003 LOGIKA Corporation. All rights reserved.

[\[about site search\]](#) [\[advertising\]](#) [\[build vertical search engine\]](#) [\[reseller program\]](#) [\[site search FAQ\]](#) [\[terms of use\]](#)

<https://web.archive.org/web/20030206151917/http://www.fusionbot.com/faqs/faq29.asp>

INTERNET ARCHIVE

WayBackMachine

http://www.fusionbot.com/faqs/faq29.asp

Go

JAN

FEB

APR

06

2002

2003

2004

About this capture

64 captures

6 Feb 2003 - 28 Jun 2017

Thursday, February 06, 2003

FusionBot®

Harness the power of seamless site search solutions

Search our Site: All of FusionBot

sitemap • search tips • faqs

LOG-INSUBSCRIBEPACKAGESFEATURESCUSTOMERSABOUT USSUPPORT

Frequently Asked Question #29

Can I prevent a particular portion / section of my page(s) from being indexed?

You can prevent a particular section of your page from being indexed by surrounding the content you wish to exclude with the `<noindex>...</noindex>` tags. The `<noindex>` tags will only work within the `<body>` of your HTML.

Please note, however, that this will only prevent the indexing of actual words on a page, but will **NOT** prevent our spider from following a specific link or links that may be included between the `<NOINDEX>` tags.

To prevent the FusionBot crawler from following a specific link or links within a page, please reference our following FAQ:

<http://www.fusionbot.com/faqs/faq47.asp>

To instruct FusionBot to omit a particular section of your content, but allow all other crawlers to continue indexing, use the `<noindexfb>...</noindexfb>` tags instead. Note the addition of "fb" to the end of the `<noindex>` tags.

Many times sites have a common footer bar or navigation bar on every page that may skew the search results, causing every page to be returned when a particular keyword is searched. The `<noindex>` or `<noindexfb>` tags will cause our spider to ignore a specific section of your page while continuing to index its remaining content.

This is different than the creation of a [robots.txt file](#) or the `META NAME="ROBOTS"` tag which will prevent an entire page or directory from being indexed.

Below is an example of the no index implementation. The tags must be implemented exactly as demonstrated to function properly. Also, after making changes to your desired pages, you must login to your account and click on the 'Request Spider' button for your changes to take effect:

```
<noindex>
This is the content you wish
to be excluded from your index
within a particular page.
</noindex>
```

Or for FusionBot only:

```
<noindexfb>
This is the content you wish
to be excluded from your index
within a particular page.
</noindexfb>
```

If you did not find an answer to your question, please return to the [Support Page](#) and fill out the form provided.



Featured Customers:



Click to Subscribe!

<< Previous FAQ

Back to FAQ List

Next FAQ >>

Copyright 1998 - 2003 LOGIKA Corporation. All rights reserved.
[\(about site search\)](#) [\(advertising\)](#) [\(build vertical search engine\)](#) [\(reseller program\)](#) [\(site search FAQ\)](#) [\(terms of use\)](#)

<https://web.archive.org/web/20051230082546/http://www.fusionbot.com:80/faqs/faq29.asp>

Frequently Asked Question #29

Can I prevent a particular portion / section of my page(s) from being indexed?

You can prevent a particular section of your page from being indexed by surrounding the content you wish to exclude with the `<noindex>...</noindex>` tags. The `<noindex>` tags will only work within the `<body>` of your HTML.

Please note, however, that this will only prevent the indexing of actual words on a page, but will **NOT** prevent our spider from following a specific link or links that may be included between the `<NOINDEX>` tags.

To prevent the FusionBot crawler from following a specific link or links within a page, please reference our following FAQ:

<http://www.fusionbot.com/faqs/faq47.asp>

To instruct FusionBot to omit a particular section of your content, but allow all other crawlers to continue indexing, use the `<noindexfb>...</noindexfb>` tags instead. Note the addition of "fb" to the end of the `<noindex>` tags.

Many times sites have a common footer bar or navigation bar on every page that may skew the search results, causing every page to be returned when a particular keyword is searched. The `<noindex>` or `<noindexfb>` tags will cause our spider to ignore a specific section of your page while continuing to index its remaining content.

This is different than the creation of a [robots.txt file](#) or the [META NAME="ROBOTS" tag](#) which will prevent an entire page or directory from being indexed.

Below is an example of the no index implementation. The tags must be implemented exactly as demonstrated to function properly. Also, after making changes to your desired pages, you must login to your account and click on the 'Request Spider' button for your changes to take effect:

```
<noindex>
This is the content you wish
to be excluded from your index
within a particular page.
</noindex>
```

Or for FusionBot only:

```
<noindexfb>
This is the content you wish
to be excluded from your index
within a particular page.
</noindexfb>
```

Alternatively, if do not wish to use pseudo-HTML like tags for instructing the FusionBot spider to ignore content, for HTML compliancy, the following comment tags, surrounding your content to ignore, are also supported by FusionBot:

```
<!--noindex begin-->...<!--noindex end-->
```

OR

```
<!--begin noindex-->...<!--end noindex-->
```

If you did not find an answer to your question, please return to the [Support Page](#) and fill out the form provided.

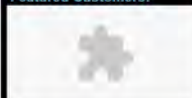
<< Previous FAQ

Back to FAQ List

Next FAQ >>



Featured Customers:



<https://web.archive.org/web/20030205145201/http://www.fusionbot.com/faqs/faq47.asp>



Search our Site: All of FusionBot GO

LOG-IN SUBSCRIBE PACKAGES FEATURES CUSTOMERS ABOUT US SUPPORT

Frequently Asked Question #47

Can I prevent a specific link on my page from being followed / indexed?

For times when you wish for FusionBot to ignore / not follow **EVERY** link on a particular page, please reference the following FAQ:

<http://www.fusionbot.com/faqs/faq8.asp>

There may also be times, however, when FusionBot **SHOULD** follow most of the links on your page, but you wish to omit a specific link or set of links from being indexed / followed within your document.

Under this scenario, you should use the **<NOFOLLOW>...</NOFOLLOW>** tags wrapped around the specific **<A HREF=...** tag in which you wish for FusionBot to ignore in your document.

Additionally, you can allow for these instructions to apply only to the FusionBot crawler by appending **"FB"** to the end of the tag name:

<NOFOLLOWFB>...</NOFOLLOWFB>

Therefore, for FusionBot to ignore the following link in your document:

Click Here

Place the following tags around the link:

<NOFOLLOW>Click Here</NOFOLLOW>

- OR -

<NOFOLLOWFB>Click Here</NOFOLLOWFB>

Please note that these links will only be ignored when our crawler is attempting to follow them from this page. However, should our crawler find these links again from within another page where the

NOFOLLOW tags are not present, they will then get indexed.

To prevent FusionBot from indexing a particular page regardless as to where it is linked from, please reference the following FAQ for implementing a robots.txt file:

<http://www.fusionbot.com/faqs/faq21.asp>

If you did not find an answer to your question, please return to the [Support Page](#) and fill out the form provided.

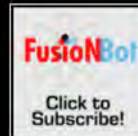
<< Previous FAQ

Back to FAQ List

Next FAQ >>



Featured Customers:



<https://web.archive.org/web/20051125113020/http://www.fusionbot.com:80/faqs/faq47.asp>

INTERNET ARCHIVE

WayBackMachine

37 captures

5 Feb 2003 - 13 Nov 2017

http://www.fusionbot.com/faqs/faq47.asp

Go

FEB

NOV

MAR

25

2004

2005

2008

About this capture

FusionBot®

Harness the power of seamless site search solutions

Site Search All of FusionBot GO

LOG-INSUBSCRIBEPACKAGESFEATURESCUSTOMERSABOUT USSUPPORT

Frequently Asked Question #47

Can I prevent a specific link on my page from being followed / indexed?

For times when you wish for FusionBot to ignore / not follow **EVERY** link on a particular page, please reference the following FAQ:

<http://www.fusionbot.com/faqs/faq8.asp>

There may also be times, however, when FusionBot **SHOULD follow** most of the links on your page, but you wish to omit a specific link or set of links from being indexed / followed within your document.

Under this scenario, you should use the **<NOFOLLOW>...</NOFOLLOW>** tags wrapped around the specific **** tag in which you wish for FusionBot to ignore in your document.

Additionally, you can allow for these instructions to apply only to the FusionBot crawler by appending **"FB"** to the end of the tag name:

<NOFOLLOWFB>...</NOFOLLOWFB>

Therefore, for FusionBot to ignore the following link in your document:

Click Here

Place the following tags around the link:

<NOFOLLOW>Click Here</NOFOLLOW>

- OR -

<NOFOLLOWFB>Click Here</NOFOLLOWFB>

Please note that these links will only be ignored when our crawler is attempting to follow them from this page. However, should our crawler find these links again from within another page where the **NOFOLLOW** tags are not present, they will then get indexed.

To prevent FusionBot from indexing a particular page regardless as to where it is linked from, please reference the following FAQ for implementing a robots.txt file:

<http://www.fusionbot.com/faqs/faq21.asp>

If you did not find an answer to your question, please return to the [Support Page](#) and fill out the form provided.

Featured Customers:





<< Previous FAQ

Back to FAQ List

Next FAQ >>

Copyright 1998 - 2005 LOGIKA Corporation. All rights reserved.

[\[about site search\]](#) [\[advertising\]](#) [\[build vertical search engine\]](#) [\[reseller program\]](#) [\[my FAQs\]](#) [\[terms of use\]](#)

<https://web.archive.org/web/20030206081202/http://www.fusionbot.com:80/whatsnew.asp>

INTERNET ARCHIVE

WayBackMachine

http://www.fusionbot.com/whatsnew.asp

Go

JUN

FEB

APR

06

2001

2003

2004

About this capture

130 captures

2 Nov 2001 - 15 Jun 2017

Thursday, February 06, 2003

FusionBot®

Harness the power of seamless site search solutions

Search our Site: All of FusionBot

sitemap • search tips • faqs

LOG-IN

SUBSCRIBE

PACKAGES

FEATURES

CUSTOMERS

ABOUT US

SUPPORT

FEATURES

- > hosted / web based
- > implementation
- > search results
- > display options
- > index controls
- > relevancy controls
- > language support
- > real-time reports
- > variety of packages
- > what's new?

What's New?

Check this page often to learn about the newest innovations implemented by the FusionBot Team, and to take full advantage of all the features your FusionBot subscription has to offer.

Feature / Enhancement	Date Implemented
1. Auto-Jump one match redirect - Using the Search Results Auto-Redirect Feature, enabling Auto-Jump will cause your visitors to be taken directly to a specific page on your site, when this page is the ONLY MATCHING DOCUMENT to a visitor's search.	1/28/2003
2. No Keywords Searched Auto-Redirect - Using the Search Results Auto-Redirect Feature, ability to redirect visitors to a specific page on your site, with a custom message, when they do not enter keywords into the search box when submitting a query.	12/20/2002
3. Download Raw Search Activity Logs - Option to download, over an extended date range (180 days), raw query logs to your computer for analysis and archival purposes, in tab, comma, or semicolon delimited format.	12/20/2002
4. Save Default Report Filter Settings - Option to automatically apply default filter settings when generating reports, preventing the re-entering of settings each time a report is requested.	12/20/2002
5. Spider Progress Viewer - After requesting a spider, you can view in real-time the spider's progress as it crawls your site. Haven't received an email confirmation, and your Index Log Report has not been updated? It's possible that the spider is still currently indexing your site. This feature will help determine whether an active spider session is in progress.	12/3/2002
6. Pass session variable values to your search results page - Silver & Gold customers can pass the value of server side session variables to their FusionBot template for re-display on their search results page.	11/26/2002
7. FusionBot Custom Meta Keywords Tag - Supplement or replace your current META Keywords tag with custom content exclusively for your FusionBot implementation.	11/28/2002
8. Reports IP Address Filter - Option to generate search reports for specific IP addresses only, omit specific IP addresses from your reports, or include all IP addresses.	11/21/2002
9. FusionBot 4.0 Launched! With this release comes a completely redesigned version of the FusionBot website. Click Here for more information.	11/8/2002
10. Create Multiple Templates for Silver and Gold subscribers. You are now able to design and implement multiple templates that allow you to control the appearance of your search results depending on where your search is being conducted from.	11/6/2002
11. Create Multiple Search Boxes for your site. You can now define multiple search partitions, each with their own unique index. This is particularly useful for keeping internal content separate from external content, yet searchable.	11/6/2002
12. Anchor Name Search - A FusionBot Exclusive! FusionBot will automatically return visitors to the specific location in your document where the match exists when your site utilizes the name attribute of the anchor tag, .	08/23/2002
13. Instruct the FusionBot spider to accept cookies when crawling your site.	08/23/2002
14. Click-Throughs Report - Discover which pages are clicked most often within your search results.	08/23/2002
15. Word Stemming for Paid Packages - Instruct FusionBot to take the STEM of your visitor's search terms to generate common variants of these words when determining matching documents.	08/23/2002
16. Common Words list for Paid Packages - Select from a standard list, or create your own custom list, of keywords that should be ignored when a visitor includes these common words in their search.	08/23/2002
17. Choose to display / disable display of icon indicating document type (Acrobat, Word, Excel)  	08/23/2002
18. For paid packages, FusionBot will now crawl and include Microsoft Word & Excel documents in your index.	08/23/2002
19. Return visitors to a specific page on your site when certain keywords or phrases are searched, rather than to the FusionBot generated search results page.	07/30/2002
20. Option to have Weekly Search Activity Reports sent to your email.	07/25/2002
21. Option to receive email notification when scheduled spider has successfully completed indexing your site.	07/25/2002
22. Emergency Email Notification Form for paid packages. A new icon is available from the Account Info tab for contacting customer support in the event of an emergency during after hours, weekends, and holidays.	05/28/2002
23. Create Custom Search Regions / Partitions for allowing visitors to search specific sections of your site.	05/16/2002
24. Change the default Mini-Portal search box text to your own custom labels.	05/16/2002



Featured Customers:



Click to Subscribe!

25.	FusionBot now follows Shockwave / Flash links for ALL packages.	05/10/2002
26.	Disable indexing of META Tag content for entire site.	04/24/2002
27.	Greater control over how sites included in your mini-portal are indexed.	04/11/2002
28.	Redirect visitors to a customized error message page on your site when no results are returned for a search.	04/10/2002
29.	Create a Robots.txt file from within your FusionBot Account via the Exclude Pages & Directories Form.	04/08/2002
30.	"Restore Previous Index" option for reverting your search results back to the previous successful crawl.	04/08/2002
31.	Specify the Default sorting option for your search results, by Rank or Date. For sites with time sensitive information, sorting the search results by Date can now be specified by default. Also, the search results give the user the ability to 're-sort' the results by Rank or Date. You may choose to disable this as an option to your users.	04/08/2002
32.	Enhanced Multi-Language Support. Select the CHARSET value from within your account rather than manually editing your FusionBot HTML code.	04/08/2002
33.	Select which language the FusionBot logo and any of the standard FusionBot generated search results text should appear in. Currently, 3 languages are available including English, Spanish, and German, with more to follow.	04/08/2002
34.	Set the value of the target frame for where your search results should appear when clicked upon from within your account rather than manually editing your FusionBot HTML code.	04/08/2002
35.	Custom setting of available Index On Demand requests or Auto-Refresh Rate frequency. For example, if your current package is adequate for most features, but you would like more Index on Demand requests to be available, or you would like to increase your Auto-Refresh Rate we can now do so, for a nominal fee, without requiring an upgrade to the next package level. Contact Customer Support for more information.	04/08/2002
36.	Silver & Gold customers benefit from enhanced customization control, and improved ease of customization with our innovative Template file & Objects syntax. Your current Header / Footer files have been automatically converted to a single template file.	02/08/2002
37.	Ability to control the elements displayed as part of the context of your search results.	02/08/2002
38.	Greater control over the options displayed as part of the search box.	02/08/2002
39.	Select from a list of pre-defined Template options for more flexibility in the appearance of your search results.	02/08/2002
40.	Custom Page Rank interface for increasing or decreasing the relevance of specific pages.	10/22/2001
41.	Facilitate the process of creating a Mini-Portal for implementing search around multiple sites.	10/15/2001
42.	Dreamweaver Extension for installation and customization of FusionBot within authoring environment.	07/24/2001
43.	Search Hits Report for viewing daily unique search terms and their corresponding number of matches, including those searches that yielded no results.	05/21/2001
44.	Suggestion Wizard for Customization Assistance.	05/01/2001
45.	Index password-protected content within your site.	04/02/2001
46.	Index Log Report for viewing status of last requested or scheduled index.	03/22/2001
47.	Index Adobe Acrobat PDF Content.	03/12/2001
48.	Instruct FusionBot NOT to index META Tag content for a specific page.	03/12/2001
49.	Adobe GoLive 5.0 Drag and Drop Implementation.	02/09/2001
50.	Prevent a particular portion / section of your page(s) from being indexed with the <noindex> - </noindex> or <noindexfb> - </noindexfb> tags.	01/16/2001
51.	Summary and Frequent Visitor Reports enhanced by adding a hyperlink to IP Address for more detailed information on who the IP Address is associated with.	01/05/2001
52.	Remove the sitemap as an option from your search box and the search results page.	12/18/2000
53.	Scoring search results. A visible 'relevancy factor' has been added to the search results. Soon you will be able to control the weight given to certain sections of your pages for even greater relevancy control.	11/27/2000
54.	Become a FusionBot Affiliate and start earning cash by referring subscribers to our service from your site!	11/08/2000
55.	Allow your visitors to search a collection of sites with a FusionBot Mini-Search Portal.	11/08/2000

<https://web.archive.org/web/20051218033318/http://www.fusionbot.com:80/whatsnew.asp>

Internet Archive

Wayback Machine

130 captures

2 Nov 2001 - 15 Jun 2017

http://www.fusionbot.com/whatsnew.asp

Go

SEP

DEC

JAN

18

2004

2005

2007

About this capture

FusionBot

Harness the power of seamless site search solutions

Site Search

All of FusionBot

GO

sitemap • search tips • faqs

LOG-IN

SUBSCRIBE

PACKAGES

FEATURES

CUSTOMERS

ABOUT US

SUPPORT

FEATURES

hosted / web based implementation search results display options index controls relevancy controls language support real-time reports variety of packages what's new?

What's New?

Check this page often to learn about the newest innovations implemented by the FusionBot Team, and to take full advantage of all the features your FusionBot subscription has to offer.

Feature / Enhancement	Date Implemented
1. Create a single search box that, depending on which menu option is selected, will provide search results from non-FusionBot related database resources / indexes	07/19/2005
2. Assign alternate search templates to specific search categories within your search box	04/20/2005
3. Import / Export your Page Relevancy Weightings. This option enables archiving (download), as well as creation, addition, modification, or deletion (upload) of your entries offline, using your preferred text editor. Customers with large lists will find this feature useful.	04/07/2005
4. Build a Staging Index - Test the accuracy of a re-crawl request before releasing your new index into production.	02/11/2005
5. Remote Re-crawl Requests - Initiate scheduled or index-on-Demand re-crawls of your site remotely, without logging into your account.	02/08/2005
6. Disable default behavior for FusionBot to automatically search for keywords within your URLs.	01/24/2005
7. Keywords Synonym List - Creation of a synonyms list enables you to specify certain words that should all be treated as synonyms, allowing your visitors to find content regardless of the specific search term they choose.	01/12/2005
8. FusionBot's search syntax now fully supports the use of boolean 'OR' searches, providing, for example, the ability to find content on a page that has either: site OR search. To specify an OR condition, the boolean indicator entered in the search box MUST be in all uppercase, OR. Additionally, the default search behavior by FusionBot is to only return pages where EVERY keyword a user enters exists on a page. You can alter this default behavior so that FusionBot will also return matching pages that have ANY of the keywords searched.	01/12/2005
9. Define search regions (partitions) via a custom FusionBot META tag - For sites that don't have content definable via directory structure, you too can now build a custom search box, or boxes, enabling your visitors to limit the scope of the search results to specific "categories" of content you define.	12/15/2004
10. Display Thumbnail Images - Insert product thumbnails / images for display as part of each matching page's result.	12/15/2004
11. FusionBot will now crawl Microsoft PowerPoint and Rich Text Format documents .	11/09/2004
12. Customize the formatting and terminology used on the 'search tips' page .	08/30/2004
13. Search Results Rollup - If you have a large number of sites listed in your mini-portal, you can choose to "rollup" the search results, so that only the first match for each distinct domain in your index is returned, followed by a link for "more results from this domain". In this manner, a site with a large number of pages won't overwhelm your visitor's search results, causing them to potentially miss important information from smaller sites.	06/14/2004
14. Sitemap Display Settings - You can now select from multiple display options for your sitemap appearance, including settings for layout and display depth.	03/16/2004
15. Click-Through Tracking Sub-Report - Once the Click-Through Report has been generated, displaying which pages are clicked most often within your search results, two sub-reports are available providing detail, for each click-through listed. These sub-reports enable you to view the total number of clicks for each page by unique visitor (IP address), or total click count for each page by unique search term conducted (i.e. What search term(s) did your visitors use when they clicked through to this page?)	03/04/2004
16. Highlighting of Search Terms within Cached Copy of Your Pages .	12/12/2003
17. Assign a Last Modified Date to your dynamic pages .	12/10/2003
18. Spell Check / Did You Mean Feature - Enable FusionBot to check the spelling of your visitor's search terms, and suggest / provide corrections.	11/21/2003
19. Integrate search results between multiple FusionBot packages .	11/07/2003
20. Multiple Search Hit Target Values - Utilizing their custom search templates, Silver & Gold customers can define multiple search hit targets to control what window a search result should display in, on a URL-by-URL basis.	10/24/2003
21. Directory Page Count Report - Curious how many pages FusionBot has found within each of the directories and sub-directories of your site or sites? The Directory Page Count report, available to Bronze and higher customers, provides a quick snapshot of page counts by directory. This can help you to refine the directories which should or should not be crawled, as well as determine whether or not visitors may benefit from the configuration of a search partition, to allow visitors the ability to search specific sections of your site exclusively, that may have a large amount of content, as displayed via this report.	10/08/2003
22. Search The Web - You can now customize your search box,	08/26/2003



Featured Customers:



Click to Subscribe!

American Airlines/Hilton, Ex. 1029, p.26

American Airlines/Hilton v. R2 Solutions

IPR2023-00690

	utilizing the Search Partitions feature , to include the ability to 'Search the Web'. Results from the Web are retrieved from LOGIKA's Galaxy.com search engine. When configuring your Search Partition, utilize the FusionBot reserved value, FB_WEB , as your Directory Path.	
23.	SSI Functionality for Search Templates - Silver and Gold customers can include remote files as part of their search templates, in order to insert dynamic content into their templates on the fly. For example, if Server Side Include (SSI) files are used for your site's menu structure or ad banner rotation, these same files can be remotely called and included within your FusionBot template.	6/1/2003
24.	Cancel Spider Request - Discovered a problem with your settings while the spider is currently crawling your site? You can cancel the currently running spider session, by selecting the 'Spider' tab, clicking on the 'Progress Viewer' link, and then clicking the 'Cancel Spider' button located at the bottom of the page. Not only will this cancel your current crawl, but your index request will be deleted from our system, resetting your number of available 'Index on Demands' to its previous level.	6/1/2003
25.	Create Search Partitions Using Wildcards - You can now include a wildcard (*) to BOTH the front AND back of your Search Partition URL entry, allowing you to specify multiple, similarly named directories throughout your entire site, which should be included as part of your region.	6/1/2003
26.	TITLE and URL Search Functionality via Search Partitions .	6/1/2003
27.	Word Stemming Applied to Page Relevance Keywords .	6/1/2003
28.	Users can now influence the position of entire domains or directories within the search results - Using the Page Relevance Weighing Feature, you can now instruct FusionBot to match partial URLs (and thus entire domains) rather than having to enter an EXACT match for influencing the position of a document. To instruct a partial match, append an asterisk (*) to the end of the URL entered in the Page URL field.	3/06/2003
29.	Auto-Jump one match redirect - Using the Search Results Auto-Redirect Feature, enabling Auto-Jump will cause your visitors to be taken directly to a specific page on your site, when this page is the ONLY MATCHING DOCUMENT to a visitor's search.	1/28/2003
30.	No Keywords Searched Auto-Redirect - Using the Search Results Auto-Redirect Feature, ability to redirect visitors to a specific page on your site, with a custom message, when they do not enter keywords into the search box when submitting a query.	12/20/2002
31.	Download Raw Search Activity Logs - Option to download, over an extended date range (180 days), raw query logs to your computer for analysis and archival purposes, in tab, comma, or semicolon delimited format.	12/20/2002
32.	Save Default Report Filter Settings - Option to automatically apply default filter settings when generating reports, preventing the re-entering of settings each time a report is requested.	12/20/2002
33.	Spider Progress Viewer - After requesting a spider, you can view in real-time the spider's progress as it crawls your site. Haven't received an email confirmation, and your Index Log Report has not been updated? It's possible that the spider is still currently indexing your site. This feature will help determine whether an active spider session is in progress.	12/3/2002
34.	Pass session variable values to your search results page - Silver & Gold customers can pass the value of server side session variables to their FusionBot template for re-display on their search results page.	11/26/2002
35.	FusionBot Custom Meta Keywords Tag - Supplement or replace your current META Keywords tag with custom content exclusively for your FusionBot implementation.	11/26/2002
36.	Reports IP Address Filter - Option to generate search reports for specific IP addresses only, omit specific IP addresses from your reports, or include all IP addresses.	11/21/2002
37.	FusionBot 4.0 Launched! With this release comes a completely redesigned version of the FusionBot website. Click Here for more information.	11/6/2002
38.	Create Multiple Templates for Silver and Gold subscribers. You are now able to design and implement multiple templates that allow you to control the appearance of your search results depending on where your search is being conducted from.	11/6/2002
39.	Create Multiple Search Boxes for your site. You can now define multiple search partitions, each with their own unique index. This is particularly useful for keeping internal content separate from external content, yet searchable.	11/6/2002
40.	Anchor Name Search - A FusionBot Exclusive! FusionBot will automatically return visitors to the specific location in your document where the match exists when your site utilizes the name attribute of the anchor tag, .	08/23/2002
41.	Instruct the FusionBot spider to accept cookies when crawling your site .	08/23/2002
42.	Click-Throughs Report - Discover which pages are clicked most often within your search results.	08/23/2002
43.	Word Stemming for Paid Packages - Instruct FusionBot to take the STEM of your visitor's search terms to generate common variants of these words when determining matching documents.	08/23/2002
44.	Common Words list for Paid Packages - Select from a standard list, or create your own custom list, of keywords that should be ignored when a visitor includes these common words in their search.	08/23/2002
45.	Choose to display / disable display of icon indicating document type (Acrobat, Word, Excel)   	08/23/2002
46.	For paid packages, FusionBot will now crawl and include Microsoft Word & Excel documents in your index.	08/23/2002
47.	Return visitors to a specific page on your site when certain keywords or phrases are searched, rather than to the FusionBot generated search results page.	07/30/2002
48.	Option to have Weekly Search Activity Reports sent to your email.	07/25/2002
49.	Option to receive email notification when scheduled spider has successfully completed indexing your site.	07/25/2002

50.	Emergency Email Notification Form for paid packages. A new icon is available from the Account Info tab for contacting customer support in the event of an emergency during after hours, weekends, and holidays.	05/28/2002
51.	Create Custom Search Regions / Partitions for allowing visitors to search specific sections of your site.	05/16/2002
52.	Change the default Mini-Portal search box text to your own custom labels.	05/16/2002
53.	FusionBot now follows Shockwave / Flash links for ALL packages.	05/10/2002
54.	Disable indexing of META Tag content for entire site.	04/24/2002
55.	Greater control over how sites included in your mini-portal are indexed.	04/11/2002
56.	Redirect visitors to a customized error message page on your site when no results are returned for a search.	04/10/2002
57.	Create a Robots.txt file from within your FusionBot Account via the Exclude Pages & Directories Form.	04/08/2002
58.	"Restore Previous Index" option for reverting your search results back to the previous successful crawl.	04/08/2002
59.	Specify the Default sorting option for your search results, by Rank or Date. For sites with time sensitive information, sorting the search results by Date can now be specified by default. Also, the search results give the user the ability to 're-sort' the results by Rank or Date. You may choose to disable this as an option to your users.	04/08/2002
60.	Enhanced Multi-Language Support. Select the CHARSET value from within your account rather than manually editing your FusionBot HTML code.	04/08/2002
61.	Select which language the FusionBot logo and any of the standard FusionBot generated search results text should appear in. Currently, 8 languages are available including English, Spanish, German, French, Italian, Dutch, Japanese, and Chinese, with more to follow.	04/08/2002
62.	Set the value of the target frame for where your search results should appear when clicked upon from within your account rather than manually editing your FusionBot HTML code.	04/08/2002
63.	Custom setting of available Index On Demand requests or Auto-Refresh Rate frequency. For example, if your current package is adequate for most features, but you would like more Index on Demand requests to be available, or you would like to increase your Auto-Refresh Rate we can now do so, for a nominal fee, without requiring an upgrade to the next package level. Contact Customer Support for more information.	04/08/2002
64.	Silver & Gold customers benefit from enhanced customization control, and improved ease of customization with our innovative Template file & Objects syntax. Your current Header / Footer files have been automatically converted to a single template file.	02/08/2002
65.	Ability to control the elements displayed as part of the context of your search results.	02/08/2002
66.	Greater control over the options displayed as part of the search box.	02/08/2002
67.	Select from a list of pre-defined Template options for more flexibility in the appearance of your search results.	02/08/2002
68.	Custom Page Rank interface for increasing or decreasing the relevance of specific pages.	10/22/2001
69.	Facilitate the process of creating a Mini-Portal for implementing search around multiple sites.	10/15/2001
70.	Dreamweaver Extension for installation and customization of FusionBot within authoring environment.	07/24/2001
71.	Search Hits Report for viewing daily unique search terms and their corresponding number of matches, including those searches that yielded no results.	05/21/2001
72.	Suggestion Wizard for Customization Assistance.	05/01/2001
73.	Index password-protected content within your site.	04/02/2001
74.	Index Log Report for viewing status of last requested or scheduled index.	03/22/2001
75.	Index Adobe Acrobat PDF Content.	03/12/2001
76.	Instruct FusionBot NOT to index META Tag content for a specific page.	03/12/2001
77.	Adobe GoLive 5.0 Drag and Drop Implementation.	02/09/2001
78.	Prevent a particular portion / section of your page(s) from being indexed with the <noindex>, </noindex> or <noindexfb>...</noindexfb> tags.	01/16/2001
79.	Summary and Frequent Visitor Reports enhanced by adding a hyperlink to IP Address for more detailed information on who the IP Address is associated with.	01/05/2001
80.	Remove the sitemap as an option from your search box and the search results page.	12/18/2000
81.	Scoring search results. A visible 'relevancy factor' has been added to the search results. Soon you will be able to control the weight given to certain sections of your pages for even greater relevancy control.	11/27/2000
82.	Allow your visitors to search a collection of sites with a FusionBot Mini-Search Portal.	11/08/2000

<https://web.archive.org/web/20051029234623/http://microformats.org/wiki/robots-exclusion>

robots-exclusion

From Microformats

Robot Exclusion Profile

[\[edit\]](#)

Table of contents [\[hide\]](#)

[1 Robot Exclusion Profile](#)

[1.1 Draft Specification 2005-06-18](#)

[1.1.1 Authors](#)

[1.1.2 Copyright](#)

[1.1.3 Patents](#)

[1.2 Abstract](#)

[1.3 Introduction](#)

[1.4 Format](#)

[1.4.1 Profile URI](#)

[1.4.2 XMDP Profile](#)

[1.5 Examples](#)

[1.6 References](#)

[1.6.1 Normative](#)

[1.6.2 Informative](#)

[1.6.3 Thanks](#)

[1.7 Issues](#)

[1.7.1 Precedence](#)

[1.7.2 Specificity](#)

[1.7.3 Keywords](#)

[1.7.4 Suitability as a microformat](#)

[\[edit\]](#)

Draft Specification 2005-06-18

Authors

[\[edit\]](#)

- [Peter Janes](http://peterjanes.ca/) (<http://peterjanes.ca/>)

Copyright

[\[edit\]](#)

This specification is © 2004-2005 by the author. However, the author intends to submit this specification to a standards body with a liberal copyright/licensing policy such as the [GMPG](http://gmpg.org/) (<http://gmpg.org/>). See the [GMPG Principles](http://gmpg.org/principles) (<http://gmpg.org/principles>) for more details. Anyone wishing to contribute to this effort MUST read those principles, especially those regarding copyright and licensing, and agree to them before contributing.

Patents

[\[edit\]](#)

The author neither holds nor intends to apply for any patents on anything required to implement this specification.

Abstract

[\[edit\]](#)

The Robot Exclusion Profile is a reworking of the Robots META tag (and less-standard extensions) as a [microformat](#).

Introduction

[\[edit\]](#)

The [Robots META tag](http://www.robotstxt.org/wc/meta-user.html) (<http://www.robotstxt.org/wc/meta-user.html>) is used to provide page-specific direction for web crawlers. While being useful in many cases, its page-specific nature means it cannot be used to restrict crawlers from indexing only certain sections of a document. Several attempts have been made to create more granular solutions through various methods but have perceived shortcomings that limit their use; the Robot Exclusion Profile defines a microformat that can be applied to any element or set of elements in a page.

Like other microformats such as [hCalendar](#), the Robot Exclusion Profile defines a set of class names that may be applied to (X)HTML elements. `class` can be applied to almost every (X)HTML element, which means that authors may be as specific or general as they wish in their application. This differs from the similarly-purposed `rel="nofollow"` attribute, which may only be applied to (and does not refer to the content of) a specific inline link. (It is interesting to note that this behaviour is entirely encompassed by the use of `class="robots-nofollow"` on the same element.) Classes are also additive, so multiple values can be specified at once, e.g. `class="robots-nofollow robots-noindex"`. For robot exclusion in particular, this allows authors to specify multiple rules for an element without adding unnecessary extra markup.

Format

[\[edit\]](#)

Profile URI

[\[edit\]](#)

<http://example.org/xmdp/robots-profile#> (obviously preliminary)

The classes defined by the Robot Exclusion Profile should be considered meaningless when the profile URI is not present in the document `<head>`'s profile attribute.

XMDP Profile

[\[edit\]](#)

```
<dl class="profile">
  <dt id="robots-nofollow">robots-nofollow</dt>
  <dd>
    Informs robots that links contained by the element are not to be followed.
  </dd>
  <dt id="robots-follow">robots-follow</dt>
  <dd>
    Informs robots that links contained by the element are to be followed.
  </dd>
  <dt id="robots-noindex">robots-noindex</dt>
  <dd>
    Informs robots that the content of the element is not to be included as part of the page.
  </dd>
  <dt id="robots-index">robots-index</dt>
  <dd>
    Informs robots that the content of the element is to be included as part of the page.
  </dd>
  <dt id="robots-noarchive">robots-noarchive</dt>
  <dd>
    Informs caching robots that the content of the element is not to be included in their cached copy.
  </dd>
  <dt id="robots-archive">robots-archive</dt>
  <dd>
    Informs caching robots that the content of the element is to be included in their cached copy.
  </dd>
</dl>
```

Examples

[\[edit\]](#)

Removing page content:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<div class="robots-noindex">There once was a man from Nantucket...</div>
<p>This page is not about <span class="robots-noindex">pornography</span>.</p>
```

Showing nofollow in conjunction with [votelinks](#), and applying it in parallel with [relnofollow](#):

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<p class="robots-nofollow">This is <a href="http://example.com/bogus">a bogus link</a>
and so is <a href="http://example.net/bogus">this</a>.</p>

<p>I don't like <a rel="nofollow" rev="vote-against" class="robots-nofollow"
href="http://example.com/disagree">this page</a>
but I do like <a rev="vote-for" href="http://example.com/agree">this one</a>.</p>
```

Preventing images from being stored by search engines, forcing them to be retrieved from the originating website:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<p></p>
```

A [more complex example](http://peterjanes.ca/2005/robots/example) (<http://peterjanes.ca/2005/robots/example>) is available which also shows how the robots metadata may be [visualized](http://tantek.com/log/2005/06.html#d03f2359) (<http://tantek.com/log/2005/06.html#d03f2359>).

References

[\[edit\]](#)

Normative

[\[edit\]](#)

- [XMDP](http://gmpg.org/xmdp/) (<http://gmpg.org/xmdp/>)
- [The Robots META Tag](http://www.robotstxt.org/wc/meta-user.html) (<http://www.robotstxt.org/wc/meta-user.html>)

Informative

[\[edit\]](#)

- [A Standard for Robot Exclusion](http://www.robotstxt.org/wc/norobots.html) (<http://www.robotstxt.org/wc/norobots.html>)
- [Googlebot Frequently Asked Questions](http://www.google.com/bot.html#noindextags) (<http://www.google.com/bot.html#noindextags>)
- [The ROBOTS META Tag](http://www.bauser.com/websnob/meta/robots.html) (<http://www.bauser.com/websnob/meta/robots.html>)
- [RelNoFollow Draft Specification](#)
- This page was contributed from the [technorati developers' wiki](http://developers.technorati.com/wiki/RobotsExclusion) (<http://developers.technorati.com/wiki/RobotsExclusion>).

Thanks

[\[edit\]](#)

- [Tantek Celik](http://tantek.com/log/) (<http://tantek.com/log/>)
- [Lachlan Hunt](http://www.lachy.id.au/) (<http://www.lachy.id.au/>)
- [Joe D'Andrea](http://www.joesapt.net/) (<http://www.joesapt.net/>)

Issues

[\[edit\]](#)

These are open issues that have been raised in various forums. The "efficacy" and "collateral damage" issues from [rel="nofollow"](#) also apply.

[\[edit\]](#)

Precedence

[[edit](#)]

- Should earlier values take precedence or later? Does `class="robots-nofollow robots-follow"` means the same as `class="robots-nofollow"` or `class="robots-follow"`?
- meta tag suggests not using conflicting or repeating directives and so does not specify precedence
- Interaction with [relnofollow](#): what does `class="robots-follow" rel="nofollow"` mean? Currently [relnofollow](#) has no profile URI defined, so the Robot Exclusion Profile takes precedence. In the future, per XMDP's [Using Multiple Profiles](#) (<http://gmpg.org/xmdp/description#multiple>), <q>the URIs in the 'profile' attribute are to be treated most significant (first) to least significant (last).</q>

Specificity

[[edit](#)]

- Does not allow control of specific UAs à la [A Standard for Robot Exclusion](#) (<http://www.robotstxt.org/wc/norobots.html>)

If it is actually necessary to control specific UAs here is an possible solution. Example:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<link rel="schema:RobotExclusion" href="http://example.org/..." />
<meta name="RobotExclusion.RobotName1" content="Foo Bot" />
<meta name="RobotExclusion.RobotName2" content="Bar Bot" />
<meta name="RobotExclusion.RobotName3" content="Evil Bot" />
</head>
<body>
<h1>Page</h1>
<p class="robots-noindex">This paragraph shouldn't be indexed by any bot.</p>
<p class="robot3-noindex">This paragraph should be indexed by every bot except "Evil Bot".</p>
<p class="robots-noindex robot1-index">This paragraph should be indexed by any bot except "Foo Bot".</p>
</div>
</body>
</html>
```

Of course it is a waste of bandwidth if there are "RobotExclusion.RobotName" meta tags on every page of a website. Thus this metatags should be stored on one page - perhaps the main page - so they can be maintained easily.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<link rel="schema:RobotExclusion" href="http://example.org/..." />
<link rel="RobotExclusion.Names" href="http://mypage.com/" />
</head>
<body>
<h1>Page</h1>
<p class="robots-noindex">This paragraph shouldn't be indexed by any bot.</p>
<p class="robot3-noindex">This paragraph should be indexed by every bot except "Evil Bot".</p>
<p class="robots-noindex robot1-index">This paragraph should be indexed by any bot except "Foo Bot".</p>
</div>
</body>
</html>
```

Keywords

[[edit](#)]

- The keywords `all` and `none` are defined by the Robots META Tag as convenience shortcuts to enable or disable the combination of `nofollow` and `noindex`, but predate Google's noarchive and should not be considered to include it. As a result, for purposes of clarity and simplicity (the [XMDP Minimalism principle](#) (<http://gmpg.org/xmdp/description#principles>)), they are not included in this version of the Robot Exclusion Profile.

Suitability as a microformat

[[edit](#)]

- Isn't the Robot Exclusion Profile designed for machines first and humans second instead of vice versa? Yes, just as much as [relnofollow](#), the deployed microformat that it's designed to replace.

Retrieved from "<http://microformats.org/wiki/robots-exclusion>"

Views

- [Article](#)
- [Discussion](#)
- [Edit](#)
- [History](#)

Personal tools

- [log in](#)

Navigation

- [Main Page](#)
- [Community portal](#)
- [Current events](#)
- [Recent changes](#)
- [Random page](#)
- [Help](#)

Search

Toolbox

- [What links here](#)
- [Related changes](#)
- [Special pages](#)



- This page was last modified 23:28, 13 Jul 2005.
- This page has been accessed 1338 times.
- [About Microformats](#)
- [Disclaimers](#)

<https://web.archive.org/web/20060104033238/http://microformats.org:80/wiki/robots-exclusion>

INTERNET ARCHIVE

WayBackMachine

263 captures

29 Oct 2005 - 6 Apr 2022

http://microformats.org/wiki/robots-exclusion

Go

OCT

JAN

FEB

04

2005

2006

2007

About this capture

Robot Exclusion Profile

Table of contents [hide]

1 Robot Exclusion Profile

1.1 Draft Specification 2005-06-18

1.1.1 Authors

1.1.2 Copyright

1.1.3 Patents

1.2 Abstract

1.3 Introduction

1.4 Format

1.4.1 Profile URI

1.4.2 XMDP Profile

1.5 Examples

1.6 References

1.6.1 Normative

1.6.2 Informative

1.6.3 Thanks

1.7 Issues

1.7.1 Precedence

1.7.2 Specificity

1.7.3 Keywords

1.7.4 Suitability as a microformat

Draft Specification 2005-06-18

Authors

■ Peter Janes

Copyright

This specification is © 2004-2005 by the author. However, the author intends to submit this specification to a standards body with a liberal copyright/licensing policy such as the [GMPG](#). See the [GMPG Principles](#) for more details. Anyone wishing to contribute to this effort MUST read those principles, especially those regarding copyright and licensing, and agree to them before contributing.

Patents

The author neither holds nor intends to apply for any patents on anything required to implement this specification.

Abstract

The Robot Exclusion Profile is a reworking of the Robots META tag (and less-standard extensions) as a [microformat](#).

Introduction

The [Robots META tag](#) is used to provide page-specific direction for web crawlers. While being useful in many cases, its page-specific nature means it cannot be used to restrict crawlers from indexing only certain sections of a document. Several attempts have been made to create more granular solutions through various methods but have perceived shortcomings that limit their use; the Robot Exclusion Profile defines a microformat that can be applied to any element or set of elements in a page.

Like other microformats such as [hCalendar](#), the Robot Exclusion Profile defines a set of class names that may be applied to (X)HTML elements. `class` can be applied to almost every (X)HTML element, which means that authors may be as specific or general as they wish in their application. This differs from the similarly-purposed `rel="nofollow"` attribute, which may only be applied to (and does not refer to the content of) a specific inline link. (It is interesting to note that this behaviour is entirely encompassed by the use of `class="robots-nofollow"` on the same element.) Classes are also additive, so multiple values can be specified at once, e.g. `class="robots-nofollow robots-noindex"`. For robot exclusion in particular, this allows authors to specify multiple rules for an element without adding unnecessary extra markup.

Format

Profile URI

<http://example.org/xmdp/robots-profile#> (obviously preliminary)

The classes defined by the Robot Exclusion Profile should be considered meaningless when the profile URI is not present in the document `<head>`'s `profile` attribute.

XMDP Profile

```
<dl class="profile">
  <dt id="robots-nofollow">robots-nofollow</dt>
  <dd>
    Informs robots that links contained by the element are not to be followed.
  </dd>
  <dt id="robots-follow">robots-follow</dt>
  <dd>
    Informs robots that links contained by the element are to be followed.
  </dd>
  <dt id="robots-noindex">robots-noindex</dt>
  <dd>
    Informs robots that the content of the element is not to be included as part of the page.
  </dd>
  <dt id="robots-index">robots-index</dt>
  <dd>
    Informs robots that the content of the element is to be included as part of the page.
  </dd>
  <dt id="robots-noarchive">robots-noarchive</dt>
  <dd>
    Informs caching robots that the content of the element is not to be included in their cached copy.
  </dd>
```

American Airlines/Hilton, Ex. 1029, p.35

American Airlines/Hilton v. R2 Solutions

IPR2023-00690

```
<dt id="robots-archive">robots-archive</dt>
<dd>
  Informs caching robots that the content of the element is to be included in their cached copy.
</dd>
</dl>
```

Examples

[\[edit\]](#)

Removing page content:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<div class="robots-noindex">There once was a man from Nantucket...</div>
<p>This page is not about <span class="robots-noindex">pornography</span>.</p>
```

Showing **nofollow** in conjunction with **votelinks**, and applying it in parallel with **relnofollow**:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<p class="robots-nofollow">This is <a href="http://example.com/bogus">a bogus link</a>
and so is <a href="http://example.net/bogus">this</a>.</p>

<p>I don't like <a rel="nofollow" rev="vote-against" class="robots-nofollow"
href="http://example.com/disagree">this page</a>
but I do like <a rev="vote-for" href="http://example.com/agree">this one</a>.</p>
```

Preventing images from being stored by search engines, forcing them to be retrieved from the originating website:

```
<head profile="http://example.org/xmdp/robots-profile#">
...
<p></p>
```

A more complex example  is available which also shows how the robots metadata may be visualized .

References

[\[edit\]](#)

Normative

[\[edit\]](#)

- XMDP 
- The Robots META Tag 

Informative

[\[edit\]](#)

- A Standard for Robot Exclusion 
- Googlebot Frequently Asked Questions 
- The ROBOTS META Tag 
- RelNoFollow Draft Specification
- This page was contributed from the technorati developers' wiki .

Thanks

[\[edit\]](#)

- Tantek Çelik 
- Lachlan Hunt 
- Joe D'Andrea 

Issues

[\[edit\]](#)

These are open issues that have been raised in various forums. The "efficacy" and "collateral damage" issues from **rel="nofollow"** also apply.

Precedence

[\[edit\]](#)

- Should earlier values take precedence or later? Does **class="robots-nofollow robots-follow"** means the same as **class="robots-nofollow"** Or **class="robots-follow"**?
- **meta** tag suggests not using conflicting or repeating directives and so does not specify precedence
- Interaction with **relnofollow**: what does **class="robots-follow" rel="nofollow"** mean? Currently **relnofollow** has no profile URI defined, so the Robot Exclusion Profile takes precedence. In the future, per XMDP's [Using Multiple Profiles](#) , <q>the URIs in the 'profile' attribute are to be treated most significant (first) to least significant (last).</q>

Specificity

[\[edit\]](#)

- Does not allow control of specific UAs à la [A Standard for Robot Exclusion](#) 

If it is actually necessary to control specific UAs here is an possible solution, Example:

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<link rel="schema.RobotExclusion" href="http://example.org/..." />
<meta name="RobotExclusion.RobotName1" content="Foo Bot" />
<meta name="RobotExclusion.RobotName2" content="Bar Bot" />
<meta name="RobotExclusion.RobotName3" content="Evil Bot" />
</head>
<body>
<h1>Page</h1>
<p class="robots-noindex">This paragraph shouldn't be indexed by any bot.</p>
<p class="robot3-noindex">This paragraph should be indexed by every bot except "Evil Bot".</p>
<p class="robots-noindex robot1-index">This paragraph should be indexed by any bot except "Foo Bot".</p>
</div>
</body>
</html>
```

Of course it is a waste of bandwidth if there are "RobotExclusion.RobotName" meta tags on every page of a website. Thus this metatags should be stored on one page - perhaps the main page - so they can be maintained easily.

```
<!DOCTYPE html PUBLIC "-//W3C//DTD XHTML 1.0 Strict//EN" "http://www.w3.org/TR/xhtml1/DTD/xhtml1-strict.dtd">
<html>
<head>
<link rel="schema.RobotExclusion" href="http://example.org/..." />
<link rel="RobotExclusion.Names" href="http://mypage.com/" />
</head>
<body>
<h1>Page</h1>
<p class="robots-noindex">This paragraph shouldn't be indexed by any bot.</p>
<p class="robot3-noindex">This paragraph should be indexed by every bot except "Evil Bot".</p>
<p class="robots-noindex robot1-index">This paragraph should be indexed by any bot except "Foo Bot".</p>
</div>
</body>
</html>
```

```
</body>
</html>
```

Keywords

[\[edit\]](#)

- The keywords `all` and `none` are defined by the Robots META Tag as convenience shortcuts to enable or disable the combination of `nofollow` and `noindex`, but predate Google's `noarchive` and should not be considered to include it. As a result, for purposes of clarity and simplicity (the [XMDP Minimalism principle](#) 🧐), they are not included in this version of the Robot Exclusion Profile.

Suitability as a microformat

[\[edit\]](#)

- Isn't the Robot Exclusion Profile designed for machines first and humans second instead of vice versa? Yes, just as much as [relnofollow](#), the deployed microformat that it's designed to replace.

This page was last modified 16:19, 5 Nov 2005.

This page has been accessed 2275 times.

[About Microformats](#)

[Disclaimers](#)

